Quasi-random spatially balanced sampling

B. L. Robertson[1] T. McDonald[2] J. A. Brown[1] C. J. Price[1]

[1]University of Canterbury, Christchurch, New Zealand

[2]Western EcoSystems Technology, Laramie, Wyoming, United States

May 6, 2021

Introduction

- Choosing an appropriate sampling design for a particular study can be difficult and there is no *best* design for all research questions.

- This choice depends on many things including the study objectives, available sampling frames and known **auxiliary variables**.

Introduction

- Choosing an appropriate sampling design for a particular study can be difficult and there is no *best* design for all research questions.

- This choice depends on many things including the study objectives, available sampling frames and known **auxiliary variables**.

- We consider estimating the population total of an environmental variable using the unbiased Horvitz-Thompson estimator (or its continuous analogue)

$$\hat{\tau} = \sum_{i=1}^{n} \frac{y_i}{\pi_i},$$

where n is the sample size, y_i is the response value at sample location $\boldsymbol{x}_i$ and $0 < \pi_i < 1$ is the inclusion probability of $\boldsymbol{x}_i$.

What is spatially-balanced sampling?

- Spatial sampling designs draw sampling locations from a study area.

What is spatially-balanced sampling?

- Spatial sampling designs draw sampling locations from a study area.
- To achieve good estimates of population characteristics, the spatial pattern of the sample should be similar to the spatial pattern of the population.

What is spatially-balanced sampling?

- Spatial sampling designs draw sampling locations from a study area.
- To achieve good estimates of population characteristics, the spatial pattern of the sample should be similar to the spatial pattern of the population.

Natural resources often exhibit spatial trends because nearby locations interact with one another and are influenced by the same set of natural and anthropogenic factors (Stevens and Olsen, 2004).

What is spatially-balanced sampling?

- Spatial sampling designs draw sampling locations from a study area.
- To achieve good estimates of population characteristics, the spatial pattern of the sample should be similar to the spatial pattern of the population.

  Natural resources often exhibit spatial trends because nearby locations interact with one another and are influenced by the same set of natural and anthropogenic factors (Stevens and Olsen, 2004).

- If nearby locations tend to be similar, there are statistical advantages to spreading the sample locations evenly over the population. A sample that is *evenly spread* is called a **spatially balanced sample**.

Spatial balance (continuous)

- Consider drawing n sample locations from a continuous resource $\Omega \subset [0,1)^2$ with $\lambda(\Omega) > 0$, where λ is the Lebesgue measure.

Spatial balance (continuous)

- Consider drawing n sample locations from a continuous resource $\Omega \subset [0,1)^2$ with $\lambda(\Omega) > 0$, where λ is the Lebesgue measure.
- Let $\pi(\boldsymbol{x}) = nf(\boldsymbol{x})$ be an inclusion density function, where $f(\boldsymbol{x}) : [0,1)^2 \rightarrow \mathbb{R}_{\geqslant 0}$ with
$$\int_{\Omega} f(\boldsymbol{x}) d\boldsymbol{x} = 1.$$

Spatial balance (continuous)

- Consider drawing n sample locations from a continuous resource $\Omega \subset [0,1)^2$ with $\lambda(\Omega) > 0$, where λ is the Lebesgue measure.

- Let $\pi(\boldsymbol{x}) = nf(\boldsymbol{x})$ be an inclusion density function, where $f(\boldsymbol{x}) : [0,1)^2 \to \mathbb{R}_{\geqslant 0}$ with
$$\int_{\Omega} f(\boldsymbol{x}) d\boldsymbol{x} = 1.$$

- A sample, $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset \Omega$, is considered spatially balanced if
$$v_i = \int_{\omega_i} \pi(\boldsymbol{x}) d\boldsymbol{x} \approx 1 \quad \text{for all } i = 1, 2, \ldots, n,$$
where ω_i is the Voronoi polygon for $\boldsymbol{x}_i$
$$\omega_i = \left\{\boldsymbol{x} \in [0,1)^2 : \|\boldsymbol{x} - \boldsymbol{x}_i\| \leqslant \|\boldsymbol{x} - \boldsymbol{x}_j\| \text{ for all } j = 1, 2, \ldots, n\right\}.$$

Spatial balance using the uniform inclusion density function

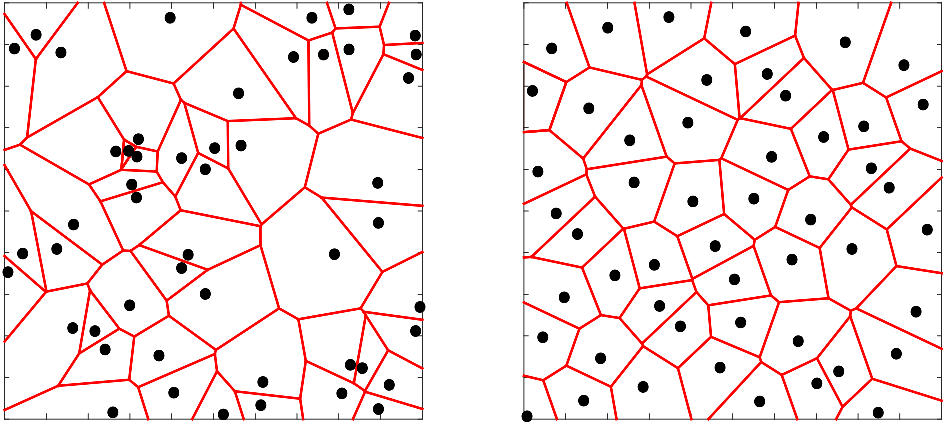


Figure: (Left) a random sample of $n = 50$ points drawn from $\Omega = [0, 1)^2$. (Right) a BAS sample of $n = 50$ points drawn from Ω. In this case, v_i is proportional to the area of ω_i (shown in red). BAS has better spatial balance than the random sample because the areas of each ω_i are more similar in size.

Spatial balance (discrete)

- Let U be a finite population of N points from $[0,1)^2$ and let $0 < \pi_i < 1$ denote the inclusion probability of $\boldsymbol{x}_i$ such that $\sum_{i=1}^{N} \pi_i = n$.

Spatial balance (discrete)

- Let U be a finite population of N points from $[0,1)^2$ and let $0 < \pi_i < 1$ denote the inclusion probability of $\boldsymbol{x}_i$ such that $\sum_{i=1}^{N} \pi_i = n$.

- A sample, $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset U$, is considered spatially balanced if

$$v_i = \sum_{\boldsymbol{x}_j \in \omega_i} \pi_j \approx 1 \quad \text{for all } i = 1, 2, \ldots, n,$$

  where ω_i is the Voronoi set for $\boldsymbol{x}_i$

$$\omega_i = \left\{\boldsymbol{x} \in U : \|\boldsymbol{x} - \boldsymbol{x}_i\| \leqslant \|\boldsymbol{x} - \boldsymbol{x}_j\| \text{ for all } j = 1, 2, \ldots, n\right\}.$$

Spatial balance for equal probability sampling

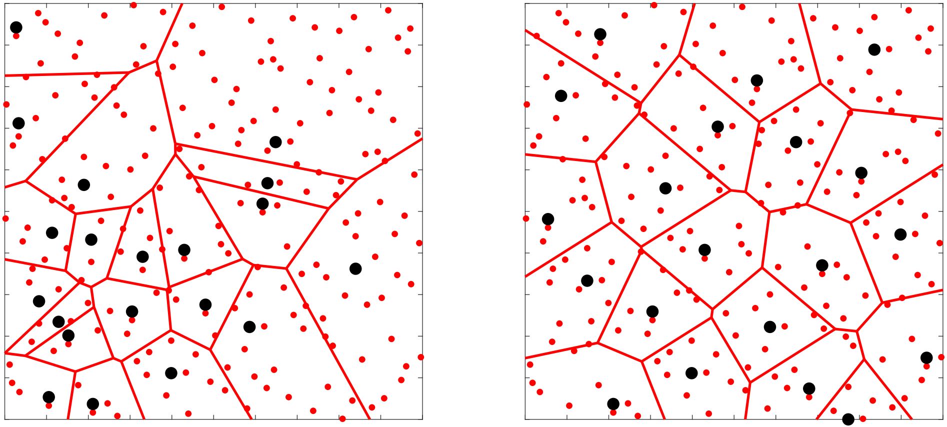


Figure: **(Left)** an SRS of $n = 20$ points drawn from $N = 200$ points. **(Right)** a HIP sample of $n = 20$ points drawn from the same 200 points. HIP has better spatial balance than SRS because the numbers of points in each ω_i are more similar.

Existing methods

- **Generalized Random Tessellation Stratified (GRTS)** (Stevens and Olsen 2004). The R packages spsurvey (Kincaid and Olsen 2015) and SDraw (McDonald 2016) can draw GRTS samples.

- **Local Pivotal Method (LPM)** (Grafström et al. 2012) and **Spatially Correlated Poisson Sampling (SCPS)** (Grafström 2011). The R package BalancedSampling (Grafström and Lisic 2016) draws LPM and SCPS samples.

- **Balanced Acceptance Sampling (BAS)** and **Halton Iterative Partitioning (HIP)** (Robertson et al. 2013,2017,2018). BAS and HIP samples can be drawn using the R package SDraw (McDonald 2016).

Psuedo-random vs. quasi-random

- In practice, randomness is introduced into sampling designs using **pseudo-random sequences**.

Psuedo-random vs. quasi-random

- In practice, randomness is introduced into sampling designs using **pseudo-random sequences**.

- These sequences are irregular, non-repetitive and designed to mimic *true random* sequences. A sequence is considered pseudo-random if it passes a series of statistical tests including distribution type, independence of successive points, runs or combinations of digits, and so on.

Psuedo-random vs. quasi-random

- In practice, randomness is introduced into sampling designs using **pseudo-random sequences**.

- These sequences are irregular, non-repetitive and designed to mimic *true random* sequences. A sequence is considered pseudo-random if it passes a series of statistical tests including distribution type, independence of successive points, runs or combinations of digits, and so on.

- In contrast, **quasi-random sequences** are not specifically designed to mimic randomness, but rather to be evenly spread over the unit box.

Quasi-random sequences

- A d-dimensional quasi-random sequence $H = \{\boldsymbol{x}_j\}_{j=1}^{n} \subset [0,1)^d$ is a sequence with the property that **for all values of n, the sequence has low discrepancy**.

Quasi-random sequences

- A d-dimensional quasi-random sequence $H = \{\boldsymbol{x}_j\}_{j=1}^{n} \subset [0,1)^d$ is a sequence with the property that **for all values of** n**, the sequence has low discrepancy**.

- The discrepancy of H is

$$D_n(H) = \sup_{B \in \mathcal{J}} \left| \frac{A_H(B)}{n} - \lambda(B) \right|, \tag{1}$$

where λ is the Lebesgue measure, $A_H(B)$ is the number of points from H in B and $\mathcal{J}$ is the set of boxes of the form

$$\{\boldsymbol{x} \in [0,1)^d : a_i \leqslant x^{(i)} < b_i\} \quad \text{with} \quad 0 \leqslant a_i < b_i < 1. \tag{2}$$

Quasi-random sequences

- A d-dimensional quasi-random sequence $H = \{\boldsymbol{x}_j\}_{j=1}^{n} \subset [0,1)^d$ is a sequence with the property that **for all values of n, the sequence has low discrepancy**.
- The discrepancy of H is

$$D_n(H) = \sup_{B \in \mathcal{J}} \left| \frac{A_H(B)}{n} - \lambda(B) \right|, \tag{1}$$

where λ is the Lebesgue measure, $A_H(B)$ is the number of points from H in B and $\mathcal{J}$ is the set of boxes of the form

$$\{\boldsymbol{x} \in [0,1)^d : a_i \leqslant x^{(i)} < b_i\} \quad \text{with} \quad 0 \leqslant a_i < b_i < 1. \tag{2}$$

Loosely speaking, a sequence is considered low-discrepancy if the fraction of points in a box is proportional to the box's size.

Quasi-random sequences

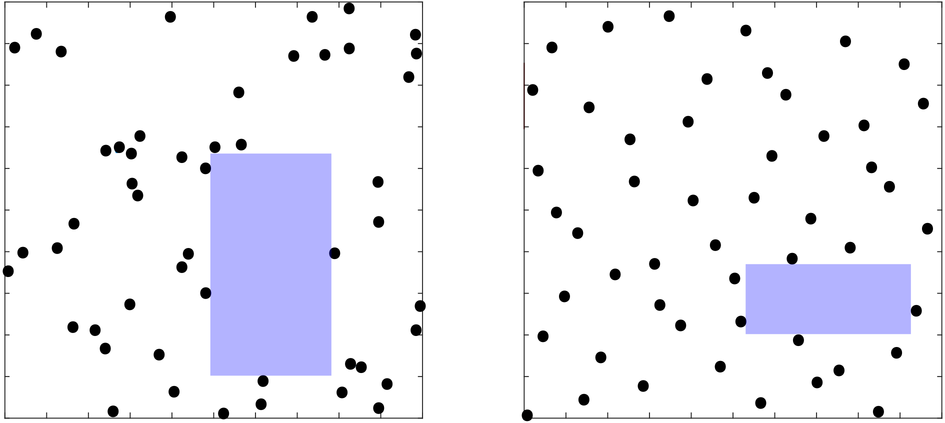


Figure: **(Left)** a random sample of $n = 50$ points drawn from $\Omega = [0, 1)^2$. **(Right)** a BAS sample of $n = 50$ points drawn from Ω. BAS is low-discrepancy, but the random sample is not.

Quasi-random sequences

- Quasi-random sequences have been used as a substitute for random numbers in many fields, including numerical integration, optimization and environmental sampling.

Quasi-random sequences

- Quasi-random sequences have been used as a substitute for random numbers in many fields, including numerical integration, optimization and environmental sampling.

- These sequences are particularly useful because they generate evenly spread points with similar spatial properties to a regular lattice. However, unlike a regular lattice, **points can be added incrementally with no clumping of points**.

Quasi-random sequences

- Quasi-random sequences have been used as a substitute for random numbers in many fields, including numerical integration, optimization and environmental sampling.
- These sequences are particularly useful because they generate evenly spread points with similar spatial properties to a regular lattice. However, unlike a regular lattice, **points can be added incrementally with no clumping of points**.
- There have been a number of different quasi-random sequences presented in the literature including the Halton (1960), the Sobol (1976) and the Faure (1982) sequences.

The Halton sequence

- The ith coordinate of the jth point in the Halton sequence $\{\boldsymbol{x}_j\}_{j=1}^{\infty} \subset [0,1)^d$ is

$$x_j^{(i)} = \sum_{p=0}^{\infty} \left\{ \left\lfloor \frac{j}{b_i^p} \right\rfloor \mod b_i \right\} \frac{1}{b_i^{p+1}},$$

where $b_1 = 2$, $b_2 = 3$ and b_i is the ith prime number and $\lfloor \cdot \rfloor$ is the floor function.

The Halton sequence

- The ith coordinate of the jth point in the Halton sequence $\{\boldsymbol{x}_j\}_{j=1}^{\infty} \subset [0,1)^d$ is

$$x_j^{(i)} = \sum_{p=0}^{\infty} \left\{ \left\lfloor \frac{j}{b_i^p} \right\rfloor \bmod b_i \right\} \frac{1}{b_i^{p+1}},$$

  where $b_1 = 2$, $b_2 = 3$ and b_i is the ith prime number and $\lfloor \cdot \rfloor$ is the floor function.

- The seventh point in the two-dimensional Halton sequence $\boldsymbol{x}_7$, for example, is

$$\begin{aligned} x_7^{(1)} &= \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{0}{16} + \tfrac{0}{32} + \dots &= \tfrac{7}{8}; \\ x_7^{(2)} &= \tfrac{1}{3} + \tfrac{2}{9} + \tfrac{0}{27} + \tfrac{0}{81} + \dots &= \tfrac{5}{9}; \\ \boldsymbol{x}_7 &= \left(\tfrac{7}{8}, \tfrac{5}{9}\right). \end{aligned}$$

The Halton sequence

- Let $B = \prod_{i=1}^{d} b_i^{J_i}$, where J_i is any non-negative integer.

The Halton sequence

- Let $B = \prod_{i=1}^{d} b_i^{J_i}$, where J_i is any non-negative integer.

- It can be shown that B **consecutive points from the Halton sequence** will have exactly **one point in each of the level B Halton boxes** defined by

$$\bigtimes_{i=1}^{d} \left[\frac{m_i}{b_i^{J_i}}, \frac{m_i + 1}{b_i^{J_i}} \right),$$

where m_i is an integer satisfying $0 \leqslant m_i < b_i^{J_i}$, for all $i = 1, \ldots, d$.

The Halton sequence

- Let $B = \prod_{i=1}^{d} b_i^{J_i}$, where J_i is any non-negative integer.

- It can be shown that B **consecutive points from the Halton sequence** will have exactly **one point in each of the level B Halton boxes** defined by

$$\bigtimes_{i=1}^{d} \left[\frac{m_i}{b_i^{J_i}}, \frac{m_i + 1}{b_i^{J_i}} \right),$$

where m_i is an integer satisfying $0 \leqslant m_i < b_i^{J_i}$, for all $i = 1, \ldots, d$.

- Hence, the Halton sequence is **quasi-periodic** (with period B) because points of the form $\boldsymbol{x}_{j+\alpha B}$ with $\alpha = 0, 1, \ldots$, are in the same level B Halton box.

Halton boxes

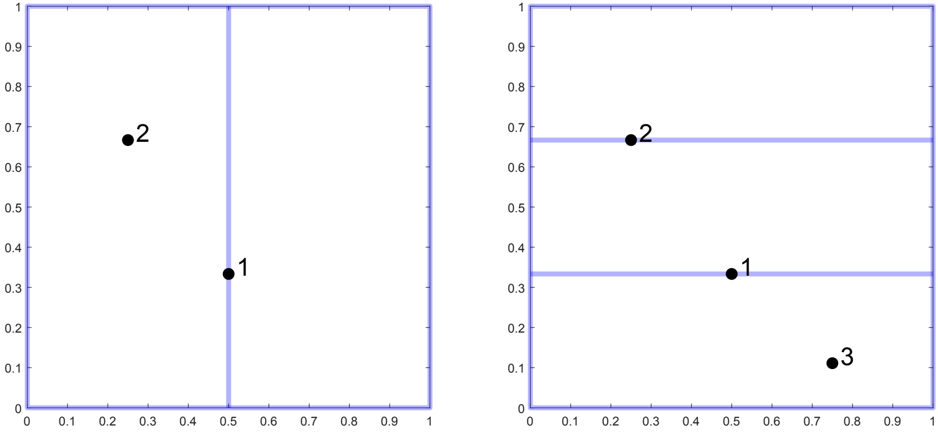


Figure: **(Left)** $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ with $B = 2^1 \times 3^0 = 2$. **(Right)** $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_3$ with $B = 2^0 \times 3^1 = 3$.

Halton boxes

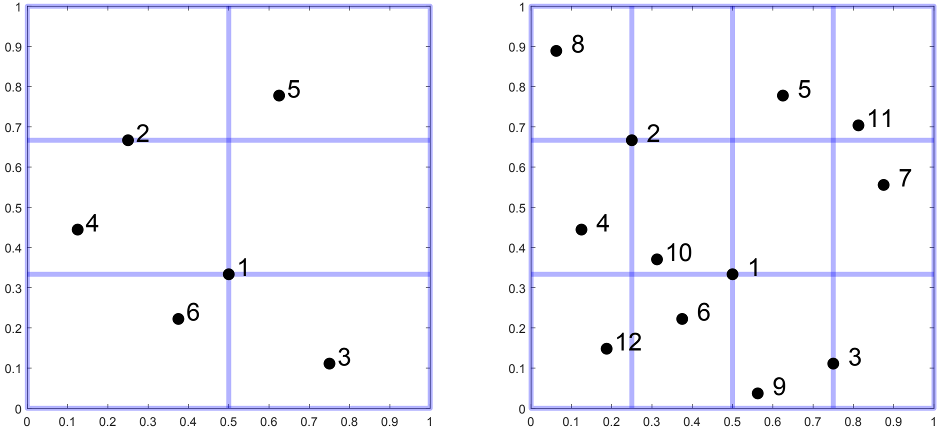
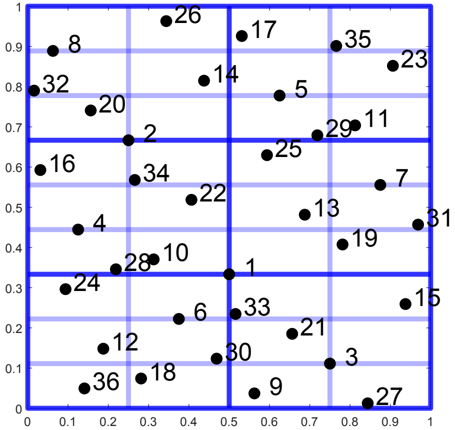


Figure: **(Left)** $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_6$ with $B = 2^1 \times 3^1 = 6$. **(Right)** $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{12}$ with $B = 2^2 \times 3^1 = 12$.



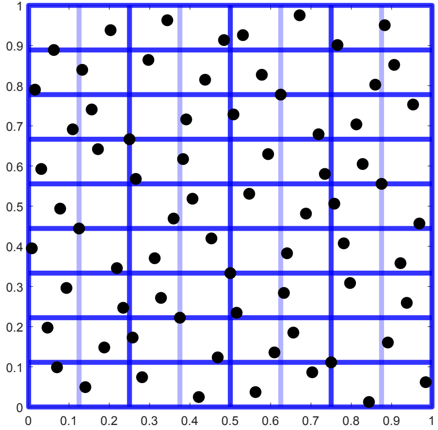
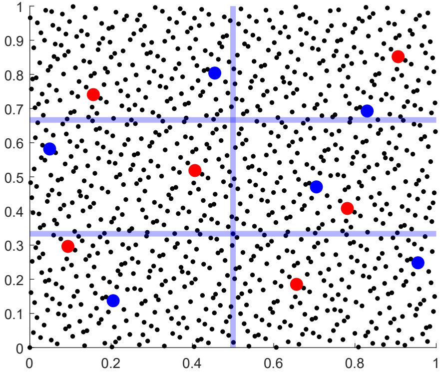


Figure: **(Left)** $x_1, \ldots, x_{36}$ with $B = 2^2 \times 3^2 = 36$. **(Right)** $x_1, \ldots, x_{72}$ with $B = 2^3 \times 3^2 = 72$.

Consecutive points in the Halton sequence



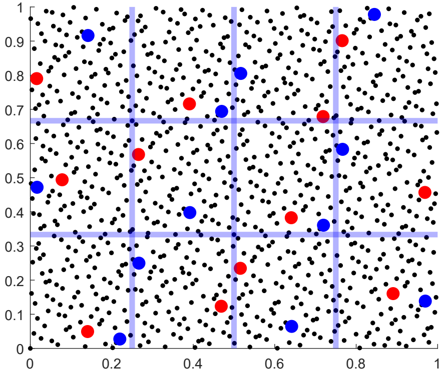


Figure: The first 2000 points from the Halton sequence. **(Left)** two sets (blue and red) of six consecutive points. **(Right)** two sets of 12 consecutive points.

Balanced Acceptance Sampling (BAS)

- BAS is a spatially balanced sampling design that draws its sample using the **random-start Halton sequence** (Wang and Hickernell 2000).

Balanced Acceptance Sampling (BAS)

- BAS is a spatially balanced sampling design that draws its sample using the **random-start Halton sequence** (Wang and Hickernell 2000).
- The ith coordinate of the jth point in this sequence $\{\boldsymbol{x}_j\}_{j=1}^{\infty} \subset [0,1)^d$ is

$$x_j^{(i)} = \sum_{p=0}^{\infty} \left\{ \left\lfloor \frac{u_i + j}{b_i^p} \right\rfloor \bmod b_i \right\} \frac{1}{b_i^{p+1}},$$

where u_i is a random non-negative integer, b_i is the ith prime number and $\lfloor \cdot \rfloor$ is the floor function.

Balanced Acceptance Sampling (BAS)

- BAS is a spatially balanced sampling design that draws its sample using the **random-start Halton sequence** (Wang and Hickernell 2000).

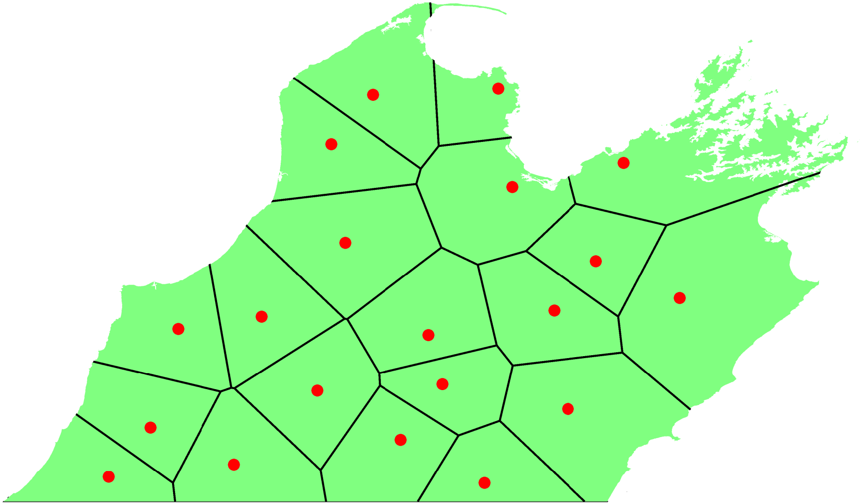
- The ith coordinate of the jth point in this sequence $\{\boldsymbol{x}_j\}_{j=1}^{\infty} \subset [0,1)^d$ is

$$x_j^{(i)} = \sum_{p=0}^{\infty} \left\{ \left\lfloor \frac{u_i + j}{b_i^p} \right\rfloor \bmod b_i \right\} \frac{1}{b_i^{p+1}},$$

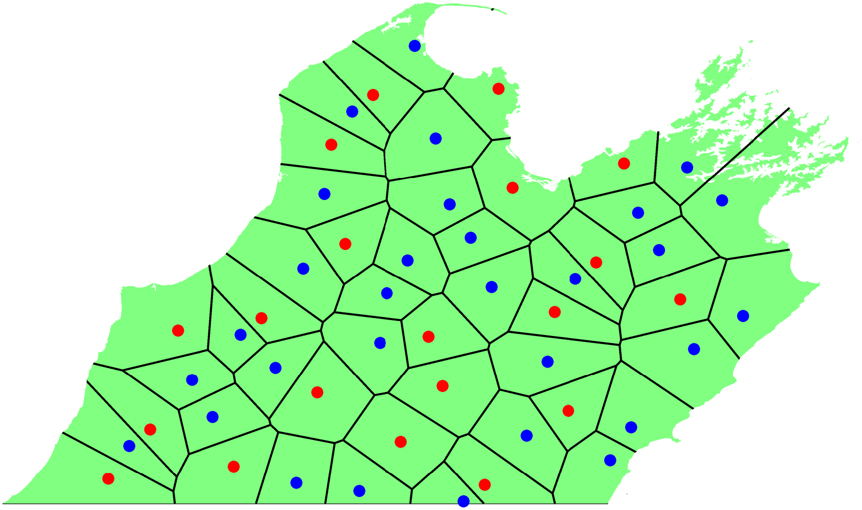
  where u_i is a random non-negative integer, b_i is the ith prime number and $\lfloor \cdot \rfloor$ is the floor function.

- Consider drawing n sample locations from a continuous resource $\Omega \subset [0,1]^2$ with positive Lebesgue measure. An **equal probability** BAS sample is simply the first n points from a random-start Halton sequence that fall within Ω (with $\boldsymbol{x}_1 \in \Omega$).

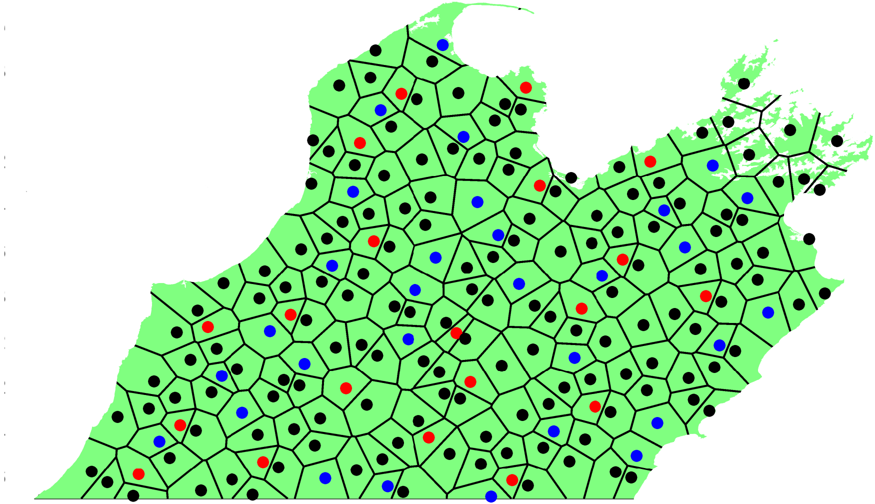
Spatial balance of BAS with $n = 20$



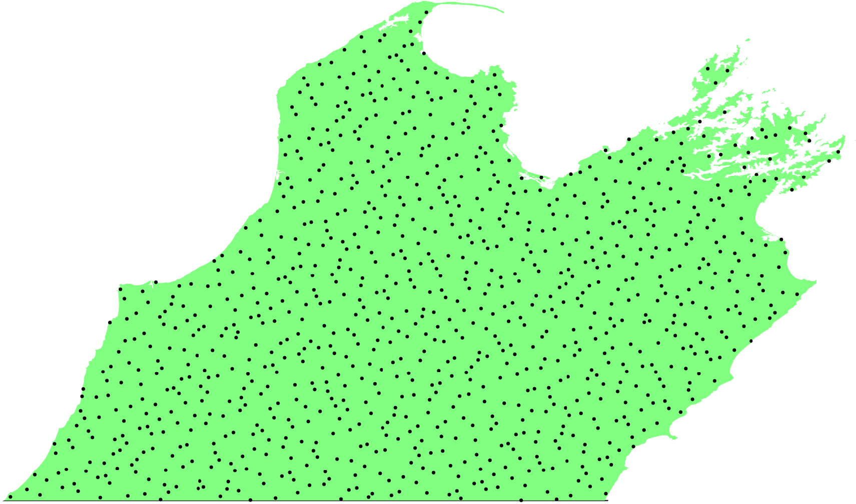
Spatial balance of BAS with $n = 50$



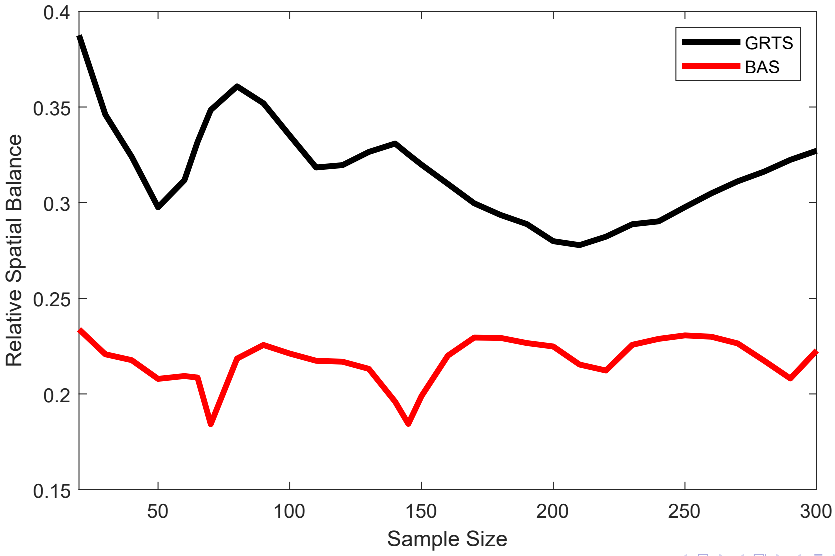
Spatial balance of BAS with $n = 200$



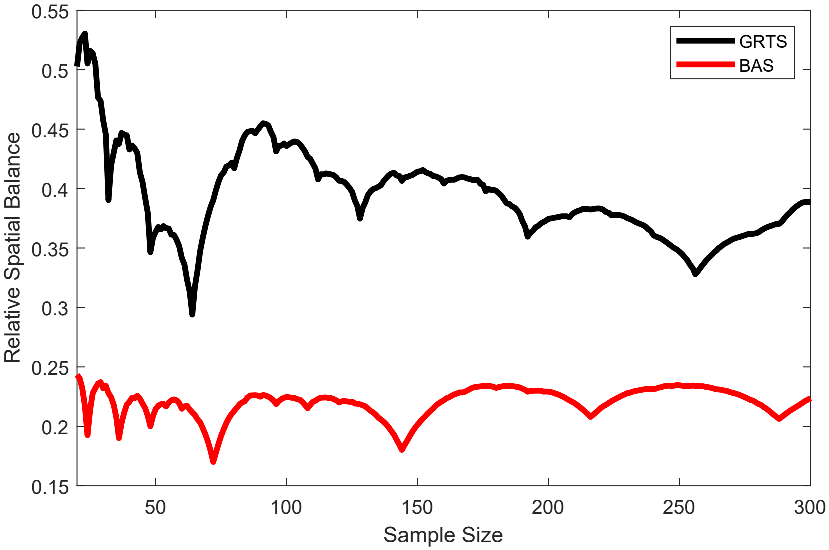
Spatial balance of BAS with $n = 1000$



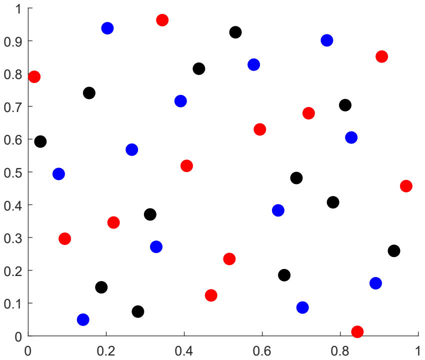
Spatial balance on $\Omega = [0,1)^2$ relative to random sampling



Spatial balance as points are added to the sample one-by-one ($\Omega = [0, 1)^2$)



Integrated monitoring between samples



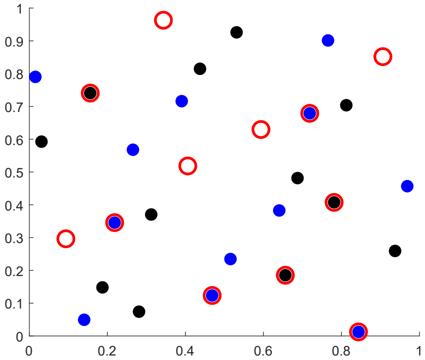
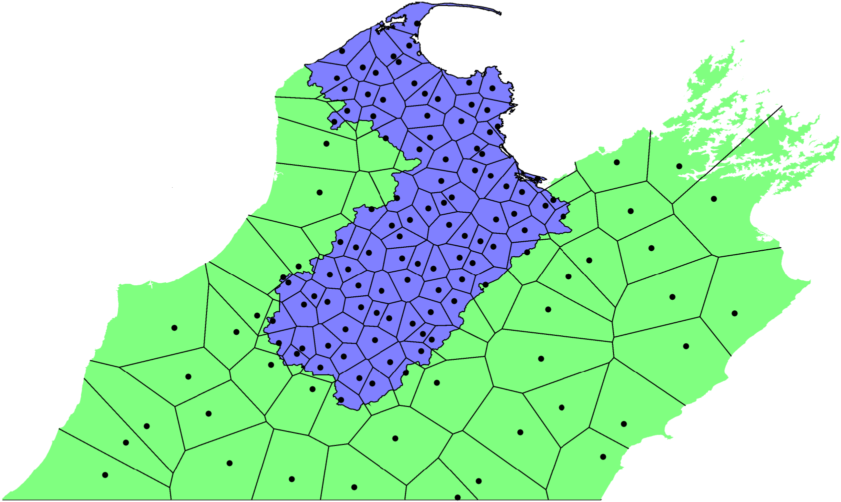


Figure: **(Left)** 36 consecutive points from the Halton sequence to define three disjoint samples with 12 points each. **(Right)** three samples of size 12 with an overlap of three.

Integrated monitoring, increasing the point density in subregions



Travelling Salesman Problem (TSP R package)

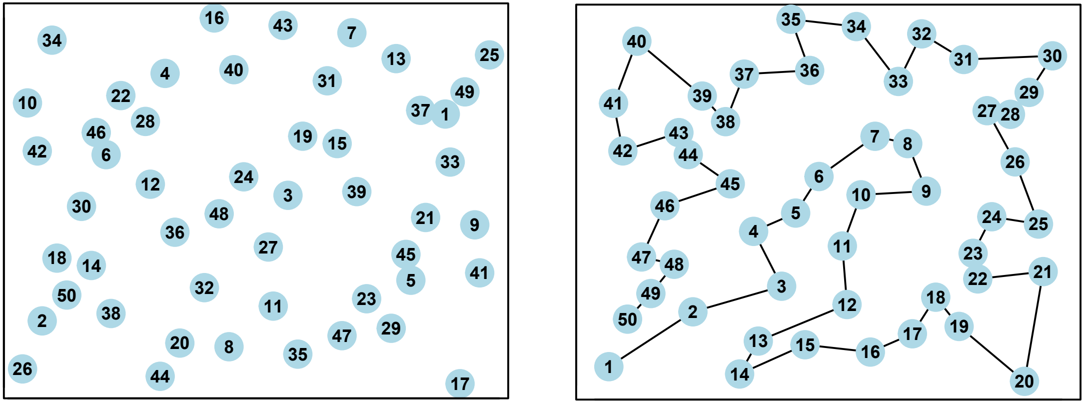


Figure: A BAS sample of $n = 50$ points. **(Left)** The ordered BAS sample such that all subsamples of consecutive points are spatially balanced. **(Right)** The BAS sample ordered for measurement efficiency.

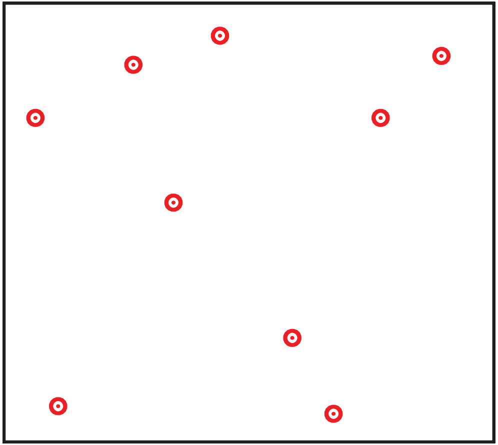
Balanced Acceptance Sampling (BAS)

- BAS is **conceptually simple**, **computationally efficient** and can adjust sample sizes dynamically to draw spatially balanced **over-samples**, **master samples** (Dam-Bates et al. 2018) or for **integrated monitoring** programmings.
- Unequal probability samples can also be drawn using a rejection sampling technique (not covered here).
- BAS can draw samples from finite populations, but it should only be used if the resource has grid structure. Otherwise BAS can be inefficient and spatial balance can be lost.

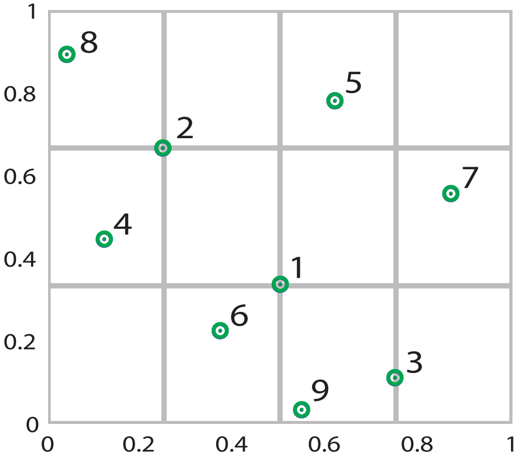
Halton Iterative Partitioning (HIP)

- Halton iterative partitioning (HIP) extends BAS to better handle point resources.
- HIP iteratively partitions a resource into $B \geqslant n$ boxes with the same nested structure as Halton boxes. These boxes are then uniquely numbered using a random-start Halton sequence of length B and the HIP sample is obtained by randomly drawing one point from each of the boxes numbered $1, 2, \ldots, n$.
- HIP is conceptually simple and computationally efficient. It uses the same ordering as the Halton sequence to ensure contiguous subsamples are spatially balanced, making it particularly useful for spatially balanced over-sampling (or a master sample) if non-target or inaccessible units are discovered.

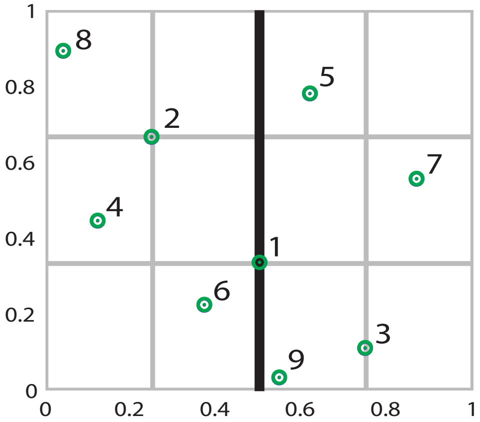
An $N = 9$ point resource

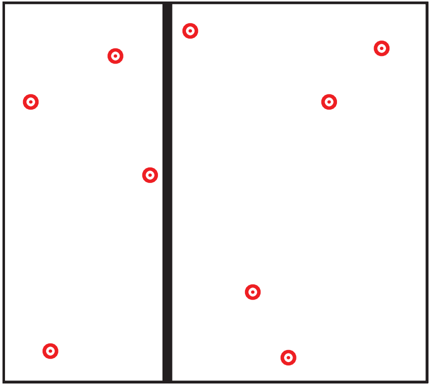


Points $x_1, \ldots, x_9$ from the Halton sequence and $B = 2^2 \times 3 = 12$ boxes

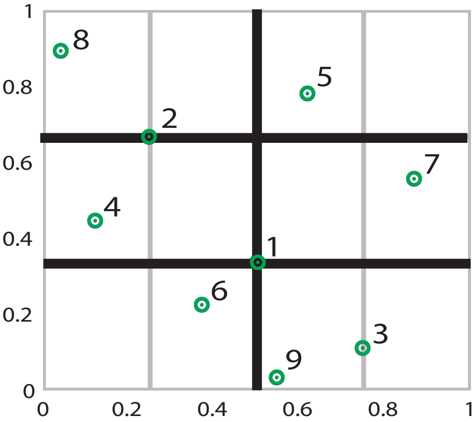


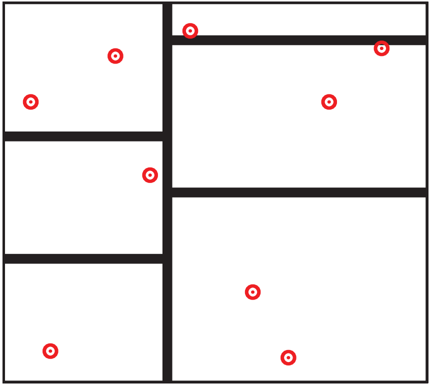
Partition with $B = 2 \times 3^0 = 2$ boxes



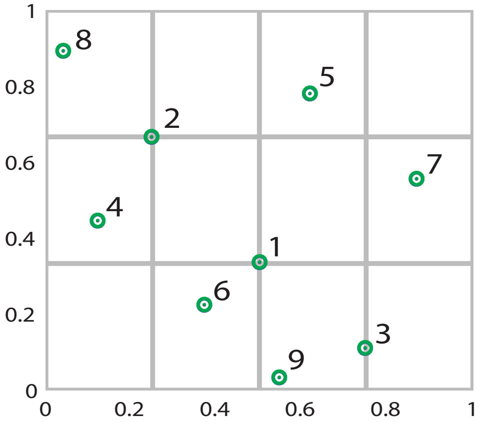


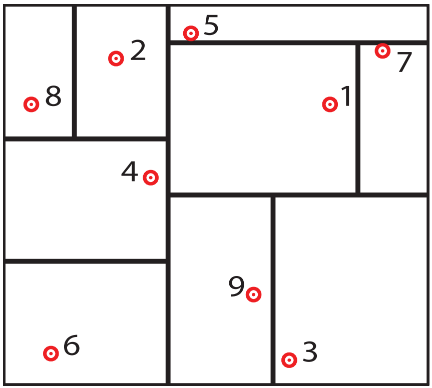
Partition with $B = 2 \times 3 = 6$ boxes



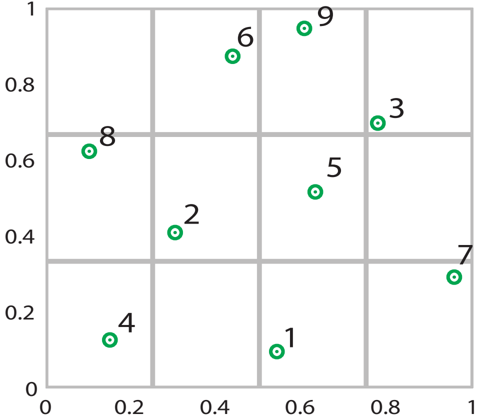


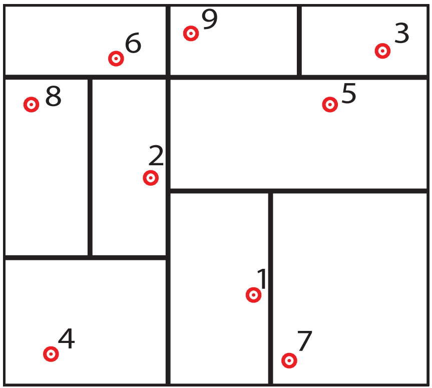
Partition with $B = 2^2 \times 3 = 12$ boxes





A random-start Halton sequence partition with $B = 2^2 \times 3 = 12$ boxes





A population of $N = 26,460$ points covering the South Island of NZ

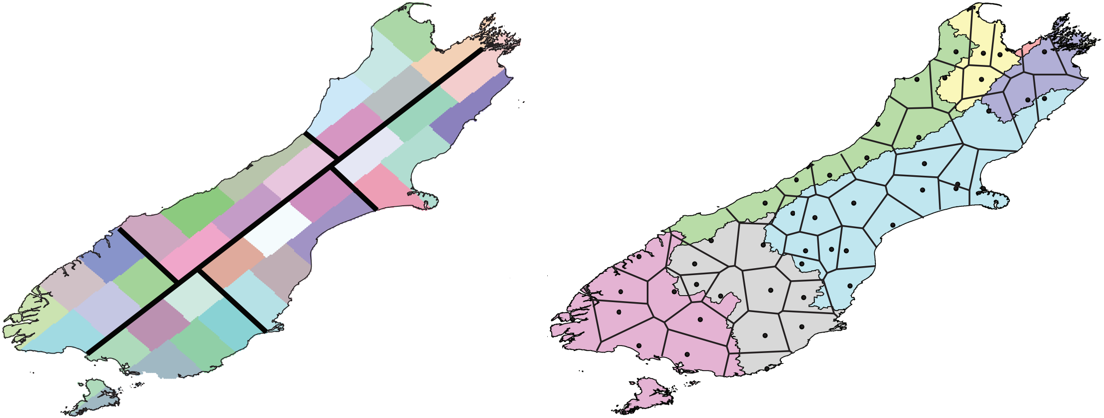


Figure: **(Left)** A HIP partition of the South Island of NZ, showing the $B = 6$ boxes (bold) and the $B = 36$ boxes (coloured). **(Right)** A HIP sample of $n = 50$ points, with geographic regions coloured.

Zooming in on the Otago region

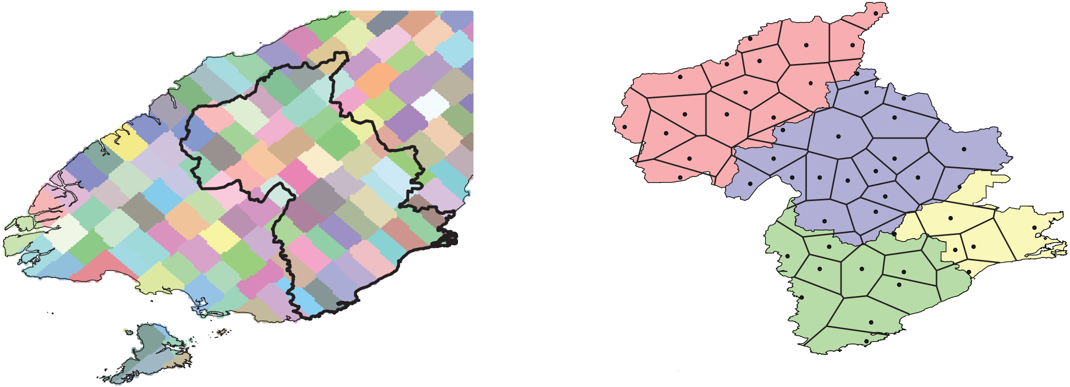


Figure: **(Left)** A zoomed in view of a HIP partition of the South Island of NZ, showing the $B = 216$ boxes (coloured). **(Right)** The $n = 50$ points from a HIP oversample that are in the Otago region, with districts coloured.

Estimation of a population total

- The Horvitz-Thompson (HT) estimator of the population total τ is

$$\widehat{\tau} = \sum_{i \in \mathcal{S}} \frac{y_i}{\pi_i},$$

where $\mathcal{S} \subset \{1, 2, \ldots, N\}$ and π_i is the inclusion probability of the ith point.

Estimation of a population total

- The Horvitz-Thompson (HT) estimator of the population total τ is

$$\hat{\tau} = \sum_{i \in \mathcal{S}} \frac{y_i}{\pi_i},$$

where $\mathcal{S} \subset \{1, 2, \ldots, N\}$ and π_i is the inclusion probability of the ith point.

- The variance of the HT estimator for a fixed sample size can be written as

$$V(\hat{\tau}) = -\frac{1}{2} \sum_{i,j} (\pi_{ij} - \pi_i \pi_j) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2,$$

and can be estimated from a sample using the Sen-Yates-Grundy estimator

$$\hat{V}(\hat{\tau}) = -\frac{1}{2} \sum_{i,j \in \mathcal{S}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2,$$

where π_{ij} is the second order inclusion probability.

Estimating the variance Horvitz-Thompson (HT) estimator

- The local mean variance estimator (Stevens and Olsen 2003) is

$$\widehat{\mathrm{V}}_{\mathrm{NBH}}(\widehat{\tau}) = \sum_{i \in \mathcal{S}} \sum_{j \in D_i} w_{ij} \left(\frac{y_j}{\pi_j} - \widehat{\tau}_{D_i} \right)^2 , \tag{3}$$

where D_i is a neighbourhood containing at least four nearest neighbours to the ith point, $\widehat{\tau}_{D_i}$ is an estimate of the population total on D_i and w_{ij} are weights.

Estimating the variance Horvitz-Thompson (HT) estimator

- The local mean variance estimator (Stevens and Olsen 2003) is

$$\widehat{\mathrm{V}}_{\mathrm{NBH}}(\widehat{\tau}) = \sum_{i \in \mathcal{S}} \sum_{j \in D_i} w_{ij} \left(\frac{y_j}{\pi_j} - \widehat{\tau}_{D_i} \right)^2, \tag{3}$$

where D_i is a neighbourhood containing at least four nearest neighbours to the ith point, $\widehat{\tau}_{D_i}$ is an estimate of the population total on D_i and w_{ij} are weights.

- Another variance estimator (Grafström and Schelin 2014) is,

$$\widehat{\mathrm{V}}_{\mathrm{G}}(\widehat{\tau}) = \frac{1}{2} \sum_{i \in \mathcal{S}} \left(\frac{y_i}{\pi_i} - \frac{y_{j_i}}{\pi_{j_i}} \right)^2, \tag{4}$$

where $j_i \in \mathcal{S}$ is the index of the nearest neighbour in the sample to the ith point.

Do spatially balanced methods actually improve on SRS?

Table: The reported values are averages using 1000 different samples, where $\mathrm{V}_{\mathrm{SIM}}$ is the empirical variance estimator. Exact values are shown for SRS, $\mathrm{V}_{\mathrm{SRS}}$.

Pop	n	SRS	GRTS		GRTS ($2n$)		HIP		
		$\mathrm{V}_{\mathrm{SRS}}$	$\mathrm{V}_{\mathrm{SIM}}$	$\widehat{\mathrm{V}}_{\mathrm{NBH}}$	$\mathrm{V}_{\mathrm{SIM}}$	$\widehat{\mathrm{V}}_{\mathrm{NBH}}$	$\mathrm{V}_{\mathrm{SIM}}$	$\widehat{\mathrm{V}}_{\mathrm{NBH}}$	$\widehat{\mathrm{V}}_{\mathrm{G}}$
	20	0.1044	0.0210	0.0282	0.0247	0.0279	**0.0132**	0.0298	0.0160
	50	0.0416	0.0041	0.0053	0.0094	0.0051	**0.0023**	0.0055	0.0026
1	100	0.0207	0.0012	0.0014	0.0041	0.0014	**0.0006**	0.0015	0.0007
	150	0.0137	0.0006	0.0006	0.0019	0.0006	**0.0002**	0.0007	0.0003
	200	0.0103	0.0004	0.0004	0.0008	0.0003	**0.0002**	0.0004	0.0002
	20	0.1812	0.0801	0.1164	0.0875	0.1122	**0.0699**	0.1162	0.1110
	50	0.0722	0.0171	0.0321	0.0285	0.0303	**0.0131**	0.0331	0.0218
2	100	0.0359	0.0071	0.0103	0.0121	0.0100	**0.0041**	0.0108	0.0057
	150	0.0238	0.0036	0.0051	0.0069	0.0049	**0.0019**	0.0053	0.0026
	200	0.0178	0.0018	0.0031	0.0036	0.0029	**0.0010**	0.0031	0.0015
	20	74.614	44.805	45.405	43.332	45.491	**31.667**	48.464	45.947
	50	29.756	9.016	13.304	12.275	12.913	**7.937**	13.590	9.843
3	100	14.803	2.873	4.431	5.480	4.316	**1.686**	4.660	2.759
	150	9.819	1.448	2.215	3.096	2.099	**0.990**	2.316	1.262
	200	7.327	0.811	1.348	1.562	1.278	**0.671**	1.368	0.714

Concluding Remarks

- Spatially balanced sampling designs are useful if nearby points are expected to have similar response values and a variety of designs have been proposed.
- BAS and HIP are spatially balanced designs that utilise the Halton sequence.
- The potential advantages of these designs over other spatially balanced designs include being **conceptually simple**, **computationally efficient**, and being able to adjust sample sizes dynamically to draw spatially balanced **over-samples**, **master samples** or for **integrated monitoring** programmings.
- Spatially balanced over-sampling is particularly useful for sampling natural resources because imperfect sampling frames and accessibility problems can result in fewer units being observed than planned. **Although the over-sampling strategy achieves desired sample sizes and is popular with field researchers, it will not eliminate the non-response or the bias of an inference.**

References

1. van Dam-Bates, P., Gansell, O. and Robertson, B. L. (2018). Using balanced acceptance sampling as a master sample for environmental surveys. *Methods in Ecology and Evolution* **9**, 1718–1726.
2. Grafström, A. and Schelin, L. (2014). How to select representative samples. *Scandinavian Journal of Statistics* **41**, 277–290.
3. Grafström, A. (2016). BalancedSampling: Balanced and Spatially Balanced Sampling. *R* package version 1.5.1. URL `https://www.antongrafstrom.se/balancedsampling`.
4. Grafström, A., Lundström, N. L. P. and Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics* **68**, 514–520.
5. Grafström, A. (2011). Spatially correlated Poisson sampling. *Journal of Statistical Planning and Inference* **142**, 36–41.
6. Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik* **2**, 84–90.
7. Kincaid, T. M. and Olsen, A. R. (2016). spsurvey: Spatial survey design and analysis. *R* package version 3.3. URL `https://CRAN.R-project.org/package=spsurvey`.
8. McDonald, T. (2020). SDraw: Spatially Balanced Samples of Spatial Objects. *R* package version 2.1.13. URL `https://github.com/tmcd82070/SDraw/wiki/SDraw`.
9. Robertson, B. L., Brown, J. A., McDonald, T., and Jaksons, P. (2013). BAS: Balanced acceptance sampling of natural resources. *Biometrics* **3**, 776-784.
10. Robertson, B. L., McDonald, T., Price, C. J., and Brown, J. A. (2017). A modification of balanced acceptance sampling. *Statistics and Probability Letters* **129**, 107-112.
11. Robertson, B. L., McDonald, T., Price, C. J., and Brown, J. A. (2018). Halton iterative partitioning: spatially balanced sampling via partitioning. *Environmental and Ecological Statistics* **25**, 305-323.
12. Robertson, B. L., van Dam-Bate, P., and Gansell, O. (2021). Halton Iterative Partitioning Master Frames. *Environmental and Ecological Statistics* (accepted).
13. Stevens, D. L., Jr. and Olsen, A. R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics* **14**, 593–610.
14. Stevens, D. L., Jr. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* **99**, 262–278.
15. Wang, X. and Hickernell, F. J. (2000). Randomized Halton sequences. *Mathematical and Computer Modelling* **32**,887–899.